LARGE GENOME SEGMENT DELETION IN *Oryza sativa* PROTOPLASTS UTILIZING CRISPR CAS9 SYNCHRONIZED CUTS

A Thesis

by

COOPER A. SVAJDA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Michael J. Thomson
Committee Members,	Russell Jessup
	Elizabeth Pierson
Intercollegiate Faculty Chair,	Dirk Hays

December 2019

Major Subject: Molecular and Environmental Plant Sciences

ABSTRACT

The newly-discovered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system has enabled rapid genome editing in plants. Moreover, CRISPR/Cas9-based editing opens up the possibility of functional characterization of large chromosomal segments, since CRISPR/Cas9 allows for large deletions targeted to specific chromosome regions. Current techniques for plant transformation, however, involve a lengthy tissue culture process for regeneration of edited plants, which creates a bottleneck when testing new targets for gene editing. This study explores the use of plant protoplasts as a testbed for rapid testing of single guide RNA (gRNA) designs for CRISPR/Cas9 gene editing in rice, using a region on rice chromosome 10 containing a nuclear-plastid DNA (NUPT) segment as a target for making a large deletion.

Plant protoplasts, which are plant cells with the cell wall removed, offer a rapid method to validate the effectiveness of CRISPR/Cas9 constructs prior to their full implementation in tissue culture. This study first optimized a plant protoplast isolation technique in rice using the Texas rice variety Presidio. Next, four different gRNAs were designed as two nested pairs, one flanking a smaller deletion (12.1 kb) and another flanking a larger deletion segment (107 kb) containing a large chloroplastic insertion. To design the gRNAs, the sequence of the Presidio genome at these locations was first obtained by sequencing PCR amplicons from the target loci. Then a ribonucleoprotein complex of Cas9 plus the gRNA for each target was used to test the gRNA efficiency *in vitro* by cutting PCR amplicons flanking each target cut site, which showed successful

editing at all four cut sites individually. Next, plasmids containing Cas9 and each pair of gRNAs were transformed into rice protoplasts, and Cas9 expression was detected after mRNA analysis, demonstrating successful expression of the plasmids within the protoplasts. Lastly, *in vivo* activity of the CRISPR system was validated for at least one of the cut sites, although the two large segment deletions were not detected in subsequent analysis of the edited protoplasts. Future efforts will be needed to further test and improve the frequency of making large chromosomal deletions before it can be widely used.

DEDICATION

I'd like to dedicate this thesis to my wife Elizabeth. If it had not been for her never-ending support of me in all that I do, nothing would have been possible to this point.

"I'm just here so I don't get fined"

-Marshawn Lynch

"Most 'Scientists' are Bottle Washers and Button Sorters"

-Lazarus Long

ACKNOWLEDGMENTS

I would like to acknowledge and thank all of my lab mates, especially Dr. Backki Kim, Oneida Ibarra, Mark Brooks and Dr. Nancy Wahl. In my time in the Crop Genome Editing Lab they taught me more about the art and science of biology that any class I have ever taken. I would like to thank my committee members, Dr. Michael Thomson, Dr. Russ Jessup, and Dr. Betsy Pierson, for their guidance and support throughout the course of this research. Finally, I would like to especially thank my past and current bosses: Dr. Thomson and Dr. Jessup were instrumental in encouraging me through the steps of graduate school as well as the steps of early adulthood.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professors Michael Thomson (Advisor) and Russell Jessup of the Department of Soil and Crop Sciences and Professor Elizabeth Pierson of the Department of Horticultural Sciences.

Funding Sources

Funding for graduate study was provided by the Texas A&M Molecular and Environmental Plant Sciences (MEPS) Rotational Fellowship, Texas A&M AgriLife Research and the H. M. Beachell Endowed Chair for International Rice Improvement held by Dr. Thomson.

NOMENCLATURE

LSD	Large Segment Deletion
RNP	Ribonucleoprotein
CRISPR	Clustered Regularly Interspaced Palindromic Repeats
TEs	Transposable Elements
CGEL	Crop Genome Editing Lab
gRNA	CRISPR/Cas9 guide RNAs
NHEJ	Non-Homologous End Joining
BS	“Big Segment”
SS	“Small Segment”
NUPTs	Nuclear Plastid DNA Segment
NUMTs	Nuclear Mitochondrial DNA Segment

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

I.1 Introduction

The future of human society is one that will be characterized by growing populations and greater pressure on an already stressed resource pool. One of the key resources that will be the deciding factor between prosperity or poverty in our future is agriculture. Fortunately for humanity, we as a species have designed an array of tools that have the potential to counteract this strain. In the world of agriculture, there are few technologies as potentially lifesaving--and as controversial--as genetic modification and genome editing. This technique offers us the potential to produce more in terms of food with less in terms of resources than ever before. It enables more consistent, higher yielding and more resilient harvests [Qaim et al., 2013].

In the past two decades a variety of different technologies have come forward as potential tools for high fidelity genome editing. These techniques, including Zinc-Finger Nucleases and TALENs, have proven to be valuable in laying the theoretical groundwork for genome editing. Amongst these technologies, CRISPR has proven to be one of the cheapest and fastest ways for performing genome edits [Carroll 2017]. This thesis will focus on the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) genome editing system. A potentially powerful facet of the world of CRISPR is the ability to remove large genome segments at once [Zhou et al., 2014]. In addition, to speed up the development of genetically modified crops with beneficial attributes, plant protoplast systems have arisen as a key testbed for the development and optimization of the CRISPR

process in target species [Priyadarshani et al., 2018] [Bottino et. al., 1981] [Zhang et. al., 2011]. By providing a cheaper and faster way to validate CRISPR constructs, testing with plant protoplast cells will prove to be an invaluable tool in the future. The objectives of this thesis are to: 1) optimize a plant protoplast isolation protocol, 2) design and validate a protocol for transformation of protoplasts with the CRISPR system, and 3) utilize the previous two steps to generate large genome segment deletions in protoplasts. The selected test case for chromosome segment deletion will target a cluster of genes from an organellar insertion in the nuclear genome of rice. In doing so it will help elucidate the differences in deletions between large and small segments, as well as look for the possible expression of chloroplast insertions in the nuclear genome of rice, all while refining the process of utilizing rice protoplasts as a testbed.

I.2 Genome Editing with CRISPR Cas9

In recent years, the new technology of CRISPR/Cas9 has enabled faster and cheaper genome editing than ever before [Carrol 2017]. CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats, which is the key feature from which the modern system was discovered. This system is a type of "bacterial immunity" against invaded viruses that enables microbes to have an "evolutionary memory" to identify and destroy phage DNA [Mojica et al., 2005]. The system that we now recognize as CRISPR in the world of biotech was first developed by Martin Jinek and others in the lab of Jennifer Doudna at UC Berkeley. In their paper, they describe how they are able to take the native CRISPR system composed of many separate functional elements, and simplify them into a user friendly two component system that we know today [Jinek et al., 2012]. This system

took advantage of a particular nuclease protein called CRISPR-associated protein 9 (Cas9). This system is comprised of a guide RNA molecule (gRNA) and the Cas9 nuclease itself.

The gRNA molecule is itself actually a manmade chimera of two separate RNAs found in bacteria. The first RNA molecule, the CRISPR RNA (crRNA), is an RNA which is complementary to the DNA sequence to be cut. The second RNA, the trans-activating crRNA (tracrRNA), is one which contains a region complementary to the tail region of the crRNA. This RNA dimer is required to interact with the nuclease, Cas9. The key step was connecting both the crRNA elements and the tracrRNA elements with a linker loop, forming a chimeric RNA we commonly call gRNA. By doing so, a streamlined two-part system was developed that reduces the overall complexity of the Cas9 system [Jinek et al., 2012].

The applicability of this technique should not be understated. Double stranded breaks are potentially fatal to cells, so cells have developed robust machinery to repair the damage quickly, primarily through non-homologous end joining (NHEJ). In the process of doing so, they typically introduce mutations. Harnessing this central process of repair and combining it with Cas9's nuclease activity is essential in most CRISPR related techniques. The most straightforward application is to create a gene knock-out. If a small insertion or deletion occurs within the coding sequence, it can cause a frameshift and premature stop codon. This can lead to a malfunctional gene product, thereby knocking out the gene function. While this has been in some ways the most studied application of CRISPR, it is by no means the only application [La Russa et al., 2015].

By supplying the cell with a desired DNA fragment, one can also use CRISPR for gene insertions and replacements in a site-specific way that was virtually impossible before. By

deleting gene suppressors up/downstream of the gene, one can actually upregulate a gene. All of this is possible with the native system. In recent years scientists have developed a plethora of systems based on the idea of removing the nuclease activity of the Cas9 protein. By doing so, they can create target-specific DNA binding proteins that can then be chimerized with any desired functional protein. This approach enables Cas9 to act like a chassis for the site-specific direction of transcriptional activators and repressors to up and down regulate target genes [La Russa et al., 2015].

The objective of this thesis is yet another unique application of CRISPR. It has been demonstrated that by performing two double stranded cuts distal to each other on the same strand of DNA, large regions of DNA can be removed at once [Cai et al., 2018] [Zhou et al., 2014]. In one study, it was demonstrated in *Glycine max* (Merr 1917) that the large segment deletions that did occur were heritable to the transformed offspring. In addition, it was observed that the larger segments that were removed saw a lower efficiency of removal than the smaller segments removed [Cai et al., 2018]. In another study, this time involving rice protoplasts, the system was demonstrated to be capable of large deletions of 170 and 245 kbp [Zhou et al., 2014]. This is a potential windfall in its biotechnological applications in its ability to knock out multiple genes or even entire quantitative trait loci (QTLs).

I.3 Organellar Insertions into the Nuclear Genome

One of the interesting quirks of our genetic code is its ability to "jump" between locations. The most well-known subset of these jumping elements are transposons [Pantzartzi et. al., 2018].

Transposons, or transposable elements (TEs), are by base pair volume some of the largest components in many eukaryotic genomes. These elements have been implicated as a powerful tool for genetic evolution, thereby elucidating why eukaryotic cells would not only tolerate them but have an active mechanism for their generation [Munoz-Lopez et al., 2010]. The process for the generation of TEs is classified as either class 1 or 2. Class 1, or retrotransposons are generated when a segment of DNA is first transcribed into RNA which is then reverse transcribed into DNA and inserted at a different genomic location. Class 2 TEs are ones which are actively generated via specific enzymes, transposases, which selectively cut the region of DNA exposing sticky ends and ligates the DNA segment into a new genomic location [Munoz-Lopez et al., 2010]. These mechanisms are able to actively move both coding and non-coding segments of the genome to different locations within the genome enabling greater diversity through chromosomal crossover.

Another, less studied class of transferred genomic DNA are composed of segments of organellar DNA which has been integrated into the nuclear genome. These segments of nuclear-mitochondrial DNAs (NUMTs) and nuclear-plastid DNAs (NUPTs) are ubiquitous in plants [Leister 2005]. The process by which these segments are integrated is still unclear. It was first postulated that these were generated from a process of RNA reverse transcription, much like class 1 transposons. However, upon further examination, no evidence for splicing or modifications, as well as segment length, indicated that the insertions were most likely the result of non-homologous end joining (NHEJ) [Leister 2005]. NHEJ is a common practice in organisms as a double stranded DNA break is potentially fatal, meaning that when a double stranded break occurs, the organism is quick to repair it, sometimes integrating any DNA segments that may be present. Earlier in the

evolutionary history of eukaryotes, the process of integrating organellar DNA into the nuclear genome also provided a fertile source for novel genetic diversity. It is estimated that up to 75% of all nuclear genes of yeast may have originated from proto mitochondria; and in the model plant *Arabidopsis*, up to 4500 genes are of plastid descent [Noutsos et al., 2005]. Despite the key role that this process played in the earlier evolution of eukaryotes, in animals this process has slowed to a point now where the insertions are significantly rarer, and in both plants and animals they contribute to primarily non-coding DNA in the nucleus when transferred, due to the lack of nuclear regulatory elements [Leister 2005].

These insertions have structural characteristics which make them useful for phylogenetic studies. Inserts of greater length (consisting of <25% of whole chloroplast genome) are typically found nearer the centromere of their respective chromosome. These larger segments typically have greater homology with their genome of origin, indicating that they are more recent in their integration [Michalovova et al., 2013]. In addition to the large segments, smaller segments are then found throughout the genome with lower levels of homology [Michalovova et al., 2013]. In some plants, such as *Arabidopsis thaliana* (Schur 1866) and *Sorghum bicolor* (Moench 1974), the occurrence of the insertions was positively correlated with the occurrence of TE's [Michalovova et al., 2013]. These patterns lead to a general conclusion for the formation and evolution of NUMTs and NUPTs. At some point a large segment of organellar DNA enters the nucleus, where it is then integrated into the genome utilizing NHEJ. The likelihood of this happening (organellar DNA present, a double stranded break, integration with NHEJ) in concert is low, hence the rarity of its occurrence. But there appear to be genomic regions prone to damage, namely the centromeres [Black et al., 2018]. So, it is then no surprise that when

these insertions occur, they typically do so around the centromere [Theuri et al., 2005]. Here the insertion lies and undergoes the typical mutations that occur to any region of the chromosome at a predictable rate. But, now that the insertion is present and in a neighborhood of typically high TE activity [Theuri et al., 2005], it can then be cut up, copied and then pasted at different sites in the genome to further its mutational evolution from its parent sequence.

By following the progression through this process, the study of organellar insertions into the nuclear genome has proven to be a powerful phylogenetic tool by being able to compare not only shared ancestry, but also the age of differentiation as well as rates of mutation [Guo et. al., 2008].

I.4 Protoplasts as a Testbed

Protoplasts are cells of organisms that possess cell walls which have had their cell walls either mechanically or enzymatically removed. They have been observed since at least the 1880's and have since become a vital part of the science of plant biology [Cocking 1972]. Because they are isolated from a parent tissue, protoplasts enable a more precise study of plant cell structure and protein localization unhindered by the "clutter" of a complete tissue much in the same way that unicellular algae played a key role in the formation of our knowledge of photosynthesis [Zallen 1993]. Over the last century, protoplasts have become a testbed for a variety of different biochemical and physiological tests that have helped us in our understanding of plant cell biology [Cocking 1972] [Jiang et. al., 2013].

Another characteristic of protoplasts is intrinsic to their cell-wall-less physiology. By being essentially an exposed plasma membrane, a variety of experiments can be performed

on them. One of the interesting possibilities is somatic fusion, or somatic hybridization. In this process, protoplasts of sometimes distant species, are fused through a variety of processes producing somatic cells with the full genetic complement of their parent lines. While the regeneration step of these protoplasts is often impossible, when it is performed it opens the door for large-scale integration of useful genes in plant pairings that are unable to produce fertile offspring via normal hybridization [Sun et al., 2004]. Seeing how the lack of a cell wall enables the introduction/mixing of whole genomes, it is no surprise that protoplasts have arisen as a powerful tool for the testing, and sometimes production, of transformed plants.

Lacking a cell wall means that DNA encoding genes of interest can be rapidly introduced rather than relying upon complex tissue culture, biolistics, agrobacterium and the like [Bottino et al., 1981]. This ability has already been utilized for a variety of gene expression, protein localization and function tests [Priyadarshani et al., 2018]. More recently, in our crop of interest rice, experiments as complex as the study of chlorophyll fluorescence and other photosynthesis-related processes have been studied in protoplast systems [Zhang et al., 2011]. For the purposes of this experiment, and the Crop Genome Editing Lab as whole, protoplasts offer a cheap and quick way to test the validity of CRISPR Cas9 constructs and gRNA effectiveness prior to the commencement of an expensive and months-long transformation of whole plants. Protoplasts are, after all, fully functioning cells that retain a large part of the genetic and epigenetic profile given to them from their host tissue [Zhang et al., 2011]. Because they not only possess the appropriate genetic machinery for the production of the CRISPR proteins, but also the particular gene

expression profiles of the target plant, they provide a tool to test the effectiveness of constructs short of executing the entire transformation and regeneration protocol.

I.5 Approach and Rationale

Bioinformatics and Plasmid Design

Rice reference genomes are currently available, and the genome browser provided by Michigan State University is the most applicable to this research (rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/). Within the user interface one is able to see the location of NUPTs and NUMTs as well as the possible gene sequences therein. In doing so, a single, large insertion of 101.2 kb was identified on the 10th chromosome of the Nipponbare reference sequence (10836654-10849000 bp). Within this insertion, a smaller gene dense region of 12.2 kb was selected for further testing (10836909-10849090 bp; Figure 1). Due to the fact that the reference sequence of an organism often differs by millions of base pairs from the actual line being tested, and the need for CRISPR gRNAs to perfectly homologous to their cut site, sequencing will be needed. Four primer pairs were designed to sequence the four flanking regions of the larger and nested smaller segments. Once these four regions have been sequenced and compared to the reference sequence for homology, four unique gRNA's will be designed for their respective cut sites.

In order to validate the effectiveness of these gRNAs on the segments of interest it is pertinent to do a controlled experiment in which all of the variables have been removed and the homology and effectiveness of the gRNA's can be validated. One of the most effective ways to do this is via the *in vitro* cleavage assay [Zhang et al., 2016]. For the *in vitro* cleavage assay the four regions that correspond to the gRNA targets will be amplified as

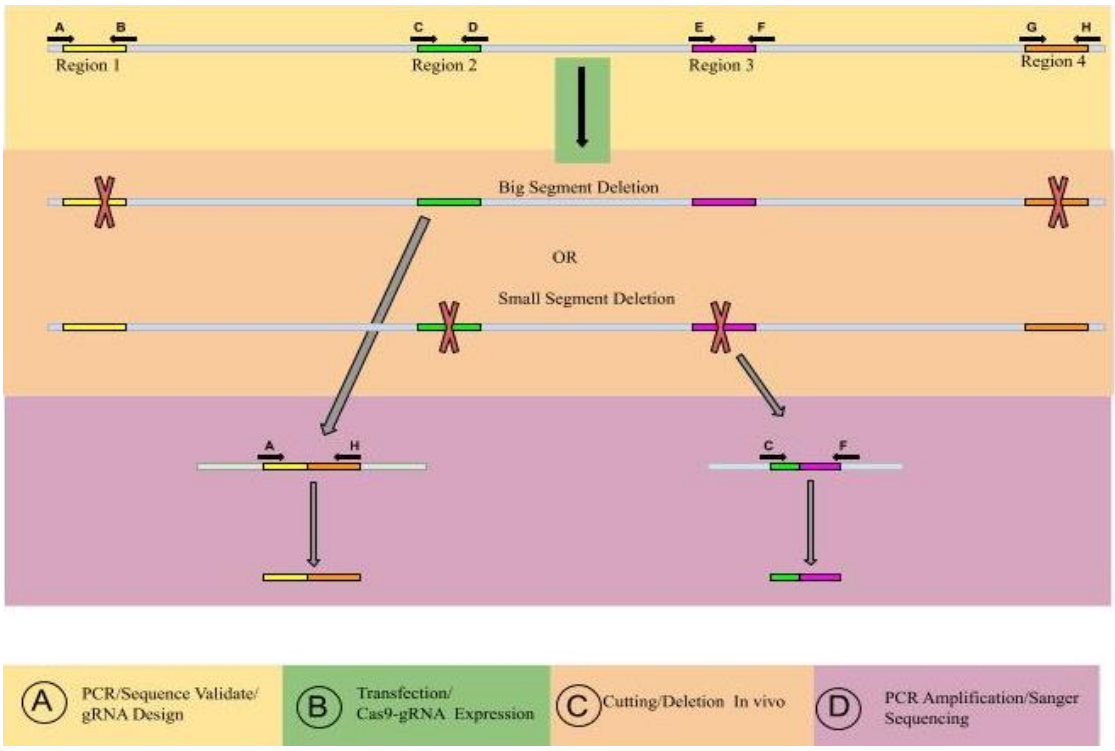


Figure 1. Graphical abstract showing the four basic phases in this experiment. First the regions 1-4 (yellow, green, purple, orange) are sequenced using PCR primers A-H. Once sequenced a gRNA for each region is designed and introduced in pairs (regions 1+4 or regions 2+3) resulting in deletions. The deletions bring their end regions together at which point the previous primer sites (A/H and C/F) are used to amplify the "repair". This PCR product is then sequenced.

PCR products from wild type Presidio [McClung 2005] rice DNA. Once the PCR products have been synthesized and run on an agarose gel to confirm product length, they will undergo a digestion with pre-assembled CRISPR Cas9 RNPs which contain the specific gRNA for each PCR product of interest [Foster et al., 2018]. After a prescribed digestion period the products will be run on a gel to confirm that the gRNA's do in fact correspond to the PCR products and that the projected cut is releasing PCR products of the desired size.

Once the gRNA's are confirmed in the *in vitro* cleavage assay, they will then be ordered in plasmid vector form for further experimentation.

The gRNA's will be ordered as polycistronic tRNA fusions on a Genscript PUC57 plasmid. The PUC57 plasmid will contain a pair of gRNAs each, one plasmid corresponding to the cut sites for the removal of the large segment, and one plasmid corresponding to the cut sites for the removal of the smaller segment. These plasmids will contain Fok1 restriction enzyme cut sites on either side of the polycistronic insertion for separation and purification. Another plasmid, PREGB32, will contain the Cas9 gene sequence and GFP sequence with a rice specific promoter. The gRNA's excised from the PUC57 will then be ligated into the sequence site of PREGB32 after a digestion with Bsa1. Once the completed plasmids (PREGB32+gRNAs for large segment, PREGB32+gRNAs for small segment) have been successfully cloned, they will then be sequenced for validation of insertion of gRNAs.

Protoplast Isolation/Transformation

Monocots, especially cereals are notoriously resistant to genetic modification. Most genotypes are recalcitrant to callus induction and regeneration and in addition *Agrobacterium ssp.* are more species specific within monocots (Hofmann 2016). This means that the process for transforming cereals for either gene knockout/insertion experiments or breeding line development is usually a time and money consuming affair. As a result, protoplasts offer a "shortcut" for the evaluation of gRNA design as well as the testing of gene function *in vivo*. The actual transformation will be performed on protoplasts isolated from 10-14-day old seedlings of the tropical *japonica* rice cultivar Presidio

[McClung 2005]. The seeds will be de-husked, washed with concentrated bleach solution then plated on a minimal media in plant tissue culture boxes. The protocol for the isolation of the protoplasts is a protocol derived from [Zhang et al., 2011]. The single step PEG transformation will then be followed with a 16-hour incubation period in the dark. The protoplast tissue samples will then have either DNA or RNA extracted using Qiagen PlantMini DNA extraction kit according to the protocols therein. For a simple validation prior to qPCR analysis, a flanking PCR will be performed. For the flanking PCR step, the two outer primers for a given region will be placed in a PCR reaction with the extracted DNA. The rationale being that the two primers would in an unaltered genome too far apart for any PCR product to be generated (100 kb or 12kb for large and small respectively). But if the whole segment deletion was in fact removed the primers will be brought into proximity enabling the production of a PCR product. If no large segment deletion is confirmed then further PCR will be required. To follow up confirmation of expression and activity of the Cas9 and gRNAs will need to be identified. For RNA analysis, the Qiagen PlantMini RNA extraction kit will be used with the protocol therein. The extracted whole RNA fraction will checked for concentration and purity using a NanoDrop spectrophotometer then be used for the production of cDNAs using the ThermoFisher cDNA generation kit with random hexamers. Once the RNA has been converted to cDNAs, a simple PCR assay will be used to validate if the plasmid was in fact transfected and expressed by looking for a confirmation of expression of the Cas9 protein. In order to see if the gRNAs are also doing their job in the cell in concert with expression of Cas9, evidence of the nucleases activity will need to be found. In order to do so, PCR products of transformed protoplast DNA containing the cut sites will be isolated and sequenced. Even

if large segment deletion is not in fact happening but nuclease activity is, then the sequencing results will show base pair mismatches on the cut site. When Cas9 cuts a region of DNA, even without large deletions, mutations occur. In the process of repairing the double stranded break the cell is prone to introduce/remove base pairs at the site of the cut. By sequencing the cut sites of DNA from transformed protoplast it will be evident, even without LSD's.

I.6 Proposed Outcome

Through this study two central questions will hopefully be answered: are protoplast in fact a reliable and attractive method for the prescreening of gRNA design before further tissue culture work, and are large segment deletions possible within the protoplast system. The first question will be answered by a positive result for the expression of Cas9 and evidence of cuts occurring at desired sites. If healthy protoplast can be isolated and transformed resulting in the expression of the Cas9 system with gRNA's, then it will be validated that protoplast are an effective tool in the rice system to validate gRNA design. The second question will be answered by the presence of the positive PCR products. If the PCR products can be generated then it will be demonstrated that not only are protoplast an effective tool for testing the Cas9 construct before tissue culture, but that they open the door for large scale removal of whole chromosomal regions (goal of greater than 10 kbp) which has a wide variety of exciting applications in plant biology.

CHAPTER II

DESIGN OF gRNAs AND PROTOPLAST ISOLATION

II.1 Synopsis

Protoplasts offer an interesting testbed for future genome editing work. By enabling a cheaper and faster way to validate the effectiveness of CRISPR Cas9 constructs prior to their full implementation in tissue culture, they can aid in increasing the timeliness and cost effectiveness of genome editing. In this chapter, the methods for the isolation and transformation of protoplast are discussed as well as the techniques used for the design of Cas9 gRNAs.

II.2 Introduction

In order to design gRNAs for the goal of LSDs, the differences in the genomes of the target species and the reference sequence (in this case that of Nipponbare) have to be accounted for. In order to do so, the region containing the cut site needs to first be isolated and sequenced. This is simply done by designing flanking primers and generating a PCR product for sequencing. So, for this experiment which will be performed on Presidio, primers were designed that would enable a PCR product to sequence and then design gRNAs using the validated sequence.

Once the sequence of the region in question is confirmed (or proven different), then the design of gRNAs is performed. Thankfully, today is a day when a whole suite of design software is available for the design of gRNAs (CRISPRdirect, Benchling Design Tool,

Synthego Design Tool and more), identification of off target effects and customization of gRNAs (Cas OFFinder).

Protoplasts are the cell wall-less cousins of either plant and bacteria cells. They are typically obtained by mechanically and then enzymatically digesting the components of the cell walls. Since their discovery in the 1880's they have proven invaluable in the study of both plants and microbes [Cocking 1972]. Due to their lack of a cell wall and their isolation as individuals from the parent tissues, protoplast offer a less resistive path for the introduction of foreign DNA into cells. This, and the fact that they can be generated in the time it takes to grow a seedling, mean that they offer a shortcut for the validation of CRISPR gRNAs as well as offering some potential for the study of genes of interest [Zallen 1993] [Priyadarshani et al., 2018]. There is a current standard protocol for the isolation of plant protoplast that was designed in the lab of Yang Zhang [Zhang et al., 2011]. One issue with this approach is that most protoplast isolation protocols are optimized for the softer, less lignified leaves of dicots. In this study a local Texas variety of rice, Presidio, was the plant of interest due to the fact that it has direct economic importance here in the state of Texas. But this line of rice is not the laboratory standard, Nipponbare. In order to enable future studies within the TAMU CGEL an optimized protocol was needed and that was the aim of this chapter. In contrast to the softer leaves of dicots, the tougher leaves of rice (even at the seedling stage) will require greater coercion to release protoplasts. In addition, due to the relative recalcitrance of rice leaves to digestion in comparison to dicots, greater amounts of plant material will be needed on the front end in order to guarantee an ample supply of protoplast for transformation.

As for transformation, one method stands above the rest in its ease of use and effectiveness. PEG, or polyethylene glycol is a simple non-reactive organic polymer that has the ability to act as a neutral osmolyte in solution. When cells immersed in a concentrated solution of PEG, they undergo several physiological changes. First due to the high osmotic pressure difference they will begin to shrink. Secondly, due to the non-polar nature of PEG, it acts much like a "reverse sterol" that makes the cell membrane both more fluid and porous. When the PEG solution also contains magnesium ions as well as DNA, the combined forces of PEG act as a conveyor that can transport the DNA across the cell membrane into the cytoplasm [Liu 2011]. By applying this method to protoplasts, their lack of a cell wall enables the easy introduction of DNA into the cytoplasm of target cells.

II.3 Materials and Methods

II.3.1 Design of gRNA's and Plasmid Vector

The region of interest is a large chloroplastic insertion on the tenth chromosome of rice (discussed further in Chapter IV). To perform this experiment, four different gRNAs would be required which correspond to four different cut sites. To design the gRNAs, the true sequence of the Presidio genome at these locations was needed. DNA was extracted from mature Presidio leaves using the CGEL modified CTAB method [Healey et al., 2014]. The NCBI PrimerDesign tool was used for the design of primers flanking the four regions based on the genome: *Oryza sativa* Japonica Group taxid: 39947. These four regions were then amplified using DreamTaq Polymerase (ThermoFischer, Waltham, Ma). Once amplified, the products were sequenced using standard Sanger sequencing. Then using the NCBI rice genome browser (www.ncbi.nlm.nih.gov/nuccore) and the

Nipponbare reference sequence NC_029265.1 was used for a MSA [Kawahara et. al., 2013]. Regions with mismatches were removed from consideration and regions with perfect homology were then used for gRNA design. Using CRISPRdirect's online tool (https://crispr.dbcls.jp) [Naito et. al., 2015], gRNAs were generated from each of the four regions and one from each was selected based on criteria of having no off target matches via Cas-OFFinder (http://www.rgenome.net/cas-offinder/) [Bae et. al., 2014].

Once the gRNAs were designed, they were ordered as a polycistronic tRNA fusion utilizing the PUC57 plasmid backbone from GenScript (GenScript, Piscataway NJ). The region containing the polycistronic insertion was excised with Fok1 restriction enzyme following protocols provided from New England Biotech (NEB, Ipswitch, MA). The binary vector pRGEB32, containing Cas9 and GFP proteins driven with an AtU6-26 promoter (Addgene, Watertown, MA), was then cut with Bsa1 utilizing protocols provided from New England Biolabs (NEB, Ipswitch, MA). The two ligation products were then ligated with T4 Ligase (NEB, Ipswitch, MA) before being inserted into heat shock competent *E. coli* DH5α (NEB, Ipswitch, MA). After colony selection on kanamycin selection plates and mass growth in liquid LB media, the plasmids were then isolated with the Qiagen MiniPrep Kit (Qiagen, Hilden, Germany).

II.3.2 Protoplast Isolation and Transformation

Seedlings for protoplast were grown from seeds that were dehulled and then soaked in 12.5% sodium hypochlorite for 30 mins before being plated on a seedling growth media (Appendix A). The seedlings were grown under full light conditions at 25 C for 11-13 days before harvest. Upon harvest the seedlings were pulled from the media, had

their roots and leaves removed and were then cut horizontally using a scalpel into pieces <1 mm in length all while in aseptic conditions. The cut pieces were immediately put into 20 mL WS1 (Appendix A) while the remainder of tissue was prepared. The WS1 was then syphoned off and pieces were immersed in 20 mL of ES (Appendix A). The dish was then wrapped in foil and placed on a shaker at 55 RPM for 4.5 hours. At the end of the 4.5-hour digestion period, the ES was removed and placed in a 50 mL falcon tube through a 100-micron filter. The tissue was then washed with 20 mL of WS2 (Appendix A) and shaken gently for 5 minutes. This material was also filtered through a 100-micron filter into a separate falcon tube. The liquids were centrifuged at 250G for 5 mins, at which point the supernatant was carefully poured off leaving the pellet of protoplast at the bottom of the tube. The pellets were resuspended in 2 mL of WS2 and collected into a single 15 mL falcon tube. This tube was centrifuged at 250G for 5 minutes and the supernatant was then poured off. The pellet was once more suspended in WS2, centrifuged and supernatant poured off. Next, the pellet was resuspended in 200 μL of MMG solution (Appendix A). This suspension was then taken to a microscope for observation.

After observation, if the protoplast were numerous, healthy and free of excessive debris, the transformation step would be undertaken. Next, 2 μg of the plasmid vector was gently mixed into the protoplasts solution. After letting this sit for five minutes at room temperature, a 40% PEG solution (Appendix A) was added to the MMG solution in a 1.1:1 ratio (220 μL PEG: 200 μL MMG). The tube was wrapped in foil and allowed to sit at room temperature for 20 minutes. At this point, the tube was centrifuged at 250G for 5 minutes. The supernatant was then removed with a pipette and the pellet was

resuspended in 1 mL of WS1 solution. The tube was then wrapped in aluminum foil and incubated at room temperature for 18 hours.

After the 18-hour incubation period, the protoplast was once again observed under the microscope in both brightfield and GFP fluorescence settings. Lastly, the protoplast was centrifuged in the tube at 1000G for five minutes before removing supernatant. The pellet was then lysed and either DNA or RNA (a whole round of isolation/transformation being used for either DNA or RNA) was extracted/purified using the Qiagen Plant Mini DNA or RNA extraction kit (Qiagen, Hilden Germany).

II.4 Results and Discussion

At this stage in the experiment, several issues were quick to arise. The original protocol for protoplast extraction (Zhang et. al., 2011) yielded too few protoplasts for a given sample of tissue with an overabundance of cellular debris. One possible hypothesis for why this is the case is that in the Zhang protocol and others, seedlings are often grown etiolated. Etiolated seedlings are grown in the dark, which leads to long, spindly seedlings with a greater amount of soft stem mass in comparison to the fibrous leaf and leaf sheath material. One of the earlier desires of this experiment was to attempt to detect gene expression from the soon-to-be-deleted insertions. These inserted genes being photosynthetic in nature, meant that to detect a difference in expression a source of light would be vital. Non-etiolated seedlings show a morphology that is more similar to normal plants meaning the above ground tissue is more green, fibrous and stouter. Owing to this changed variable it was possible that the treatment was not severe enough or too many protoplasts were lost in the Zhang protocol's various wash/transfer stages. To remedy

this, the amount of cellulase and macroenzymes per volume of enzyme solution (ES) was increased by 50%. In addition, the duration of digestion was increased from 3 to 4.5 hours at a 55 vs 50 RPM.

In addition to this, many protocols call for the discarding of the enzyme solution after digestion due to the amount of cellular debris present. The obvious downside to this is that a great majority of the liberated cells are in suspension at this point meaning that the removal of the solution results in the removal of the cells. The number of washes is often 4 or greater in most protocols and this also leads to lost protoplasts. And lastly, after transformation, most protoplast are then placed in 6 well plates to enable greater gas exchange to enhance the survival of protoplasts during incubation. Once placed in the 6 well plates, the protoplasts tend to affix to the bottom and walls of plate and are very resistant to being rinsed off for further use. To remedy these issues, all the enzyme solution and following wash solution were centrifuged together to maximize protoplasts. Next, the number of wash steps was reduced to two limiting the number of times that the protoplasts could be lost in supernatant removal. Lastly, rather than placing the incubating protoplasts in the six well plate, they were instead placed in a 2 mL microcentrifuge tube that was laid on its side to enable the necessary gas exchange. All of these modifications meant that the lower number of liberated protoplasts was a non-issue due to the fact that a higher proportion of them made it to the final stages of the experiment.

One of the key goals at this stage of the experiment was to find evidence of GFP expression. The pRGEB32 binary vector contains both the Cas9 and GFP proteins under the same promoter that was meant to make confirmation of transformation and expression

as easy as seeing GFP fluorescence. Despite this, no signal could be detected that was strong and distinct enough to call GFP fluorescence (Figure 2).

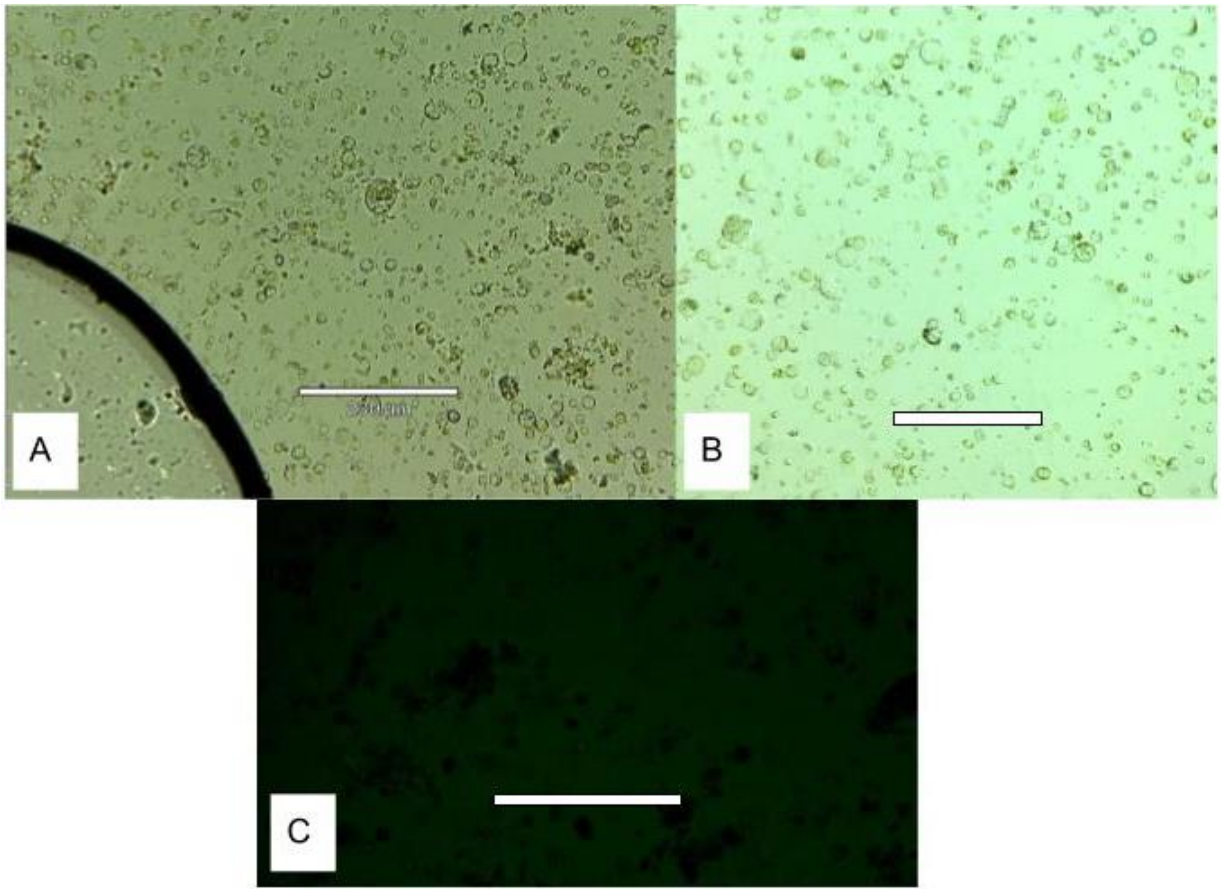


Figure 2. Collage depicting protoplasts digestion process, including images of protoplasts under brightfield and fluorescence microscope. Above figure is of brightfield and GFP images of protoplast at magnification. Image (A) shows protoplasts directly after extraction showing large numbers of healthy cells, but also large amounts of burst cells and cellular debris. Image (B) shows protoplast after 18-hour incubation. Note the lower density of cells indicating that many cells were lysed and then degraded due to PPG treatment and incubation. Image (C) shows GFP images of protoplasts (530/460 nm emission / excitation) from same sample as image (B). Some protoplast had slightly higher levels of brightness but none that was bright enough to be considered true GFP fluorescence. (Scale bars A=230 µM, B=120 µM, C=120 µM)

As will be seen in later chapters, transformation and expression with the vector was confirmed using other means. This leaves a question, if transformation and expression was occurring, why was it not observable via GFP? One simple reason could be that while the plasmid was being expressed at levels high enough for nuclease activity, it simply was not producing enough GFP protein for the fluorescence to be observable. This is the inferred cause of the lack of fluorescence. In the period after isolation, protoplast cells are highly stressed both due to the extreme stresses of cell wall degradation, lack of nutrient source and osmotic stress (notably from PEG immersion). Also, the cells were grown and developed in a tissue specific environment in which both environmental and hormonal signals were carefully maintained. Once removed from this set of "guiding influences" the cells often undergo a certain level of reduced metabolism and activity in order to help survival [Cocking 1972]. So while protein expression is certainly still occurring, it is likely that the expression is not occurring at a high enough level for the GFP produced to be visible at any exposure setting on the microscope.

II.5 Conclusion

The key findings of this portion of the experiment are that protoplasts are in fact quick and easy to generate in addition to being amenable to genetic transformation for CRISPR studies. Even though rice are monocots with a slightly more fibrous texture, their seedlings offer a readily available source of tissues that can greatly reduce the time and monetary requirements of vector construction validation.

CHAPTER III

CONFIRMATION OF gRNA ACTIVITY AND *IN VIVO* ACTIVITY

III.1 Synopsis

The primary purpose of this study was to create a streamlined protocol for the use of protoplasts as a screening tool for the designs of gRNA/Cas9 vectors, in the process testing the limits of large segment deletions. But, in order to optimize the protocol, variables need to be controlled for in order to have a clearer picture of what is occurring. The possibility of their not being a large segment deletion in this experiment necessitated alternative ways to validate the effectiveness of both the gRNA's, as well as the effectiveness of the transformation protocol and expression/activity of expressed proteins and gRNA's. The aim of this chapter is to discuss the ways in which this was accomplished.

III.2 Introduction

When optimizing a protocol or performing any experiment in general, it is necessary to isolate and eliminate variables. This means that as one walks through the steps of an experiment they are controlling for other factors so the result that is seen can be attributed to the expected cause. In the case of this experiment the final goal is a large segment deletion from the genome of protoplasts *in vivo*. In order for this to happen, though, a variety of other key steps need to be working in order. Before the ability to remove a whole region is confirmed, it needs to be known if the plasmid is in fact making into the protoplasts, it is being expressed and that the Cas9 nuclease is cutting and repair

mechanisms are occurring as expected. Add to that the effectiveness of the gRNA itself and we were left with four key variables, any one of which could derail the experiment if not performing as expected.

One of the quickest and clearest ways to test for the efficacy of the gRNAs on the regions of interest is the *in vitro* cleavage assay [Zhang et al., 2016]. In this assay, synthesized gRNAs are combined with active Cas9 proteins in order to form what are known as RNPs, or ribonucleoproteins. These are essentially what the cells produce themselves once transformed, except that they are being generated and used *in vitro* as opposed to *in vivo*. The process of making the RNPs is as simple as mixing the separate components and allowing them to combine, and then placing them in the presence of their intended targets. By doing so, a PCR product of a known size containing the cut site can be digested and ran on a gel. Once run on the gel, the known length PCR fragment will then appear as two separate bands of known length based on where the cut site was located in the original fragment [Foster et al., 2018]. Due to the acellular nature of this process there is no NHEJ occurring, leading to clear, easily identifiable bands indicating success.

The next key variables are the transformation and expression steps. As was alluded to in Chapter II, the intended screening step was the visualization of GFP, but due to the lack of fluorescence that was visible, another way was needed. The simplest way to validate this was merely detecting presence of Cas9 mRNA within the extracted RNA fraction from protoplasts. By detecting the presence of Cas9 mRNA via a PCR reaction containing primers that are homologous to the coding sequence of Cas9, it can be validated that the plasmid was both transferred to the cytoplasm and expressed.

The last variable was the *in vivo* activity of the nuclease in concert with the attached gRNAs. Whereas a large segment deletion is theoretically more obvious for reasons that will be discussed in the next chapter, proving the effectiveness of the system on an individual region basis is somewhat trickier. The nature of this process means that rather than looking for large missing DNA segments, one has to look for small DNA alterations [Jinek et. al., 2012]. When a single Cas9 mediated double strand break occurs in the absence of other cuts, it is quickly annealed using the cells endogenous repair mechanisms, most often NHEJ. This is confounded by the fact that different nuclease proteins can cause cuts of different geometry, ranging from the standard blunt end cut to the "sticky end" overhang cut. These cuts can be detected via different methods, such as high-resolution melting analysis and T7 assay [Denbow et. al., 2018] [Kleinstiver et. al., 2016]. But for this study Sanger sequencing was chosen as the best method. In doing so the DNA from transformed protoplasts was run through PCR to generate products containing the individual cut sites and then TA cloned. Once cloned and then sequenced, it is possible to detect small differences at the cut site such as single base pair insertions and/or deletions.

III.3 Materials and Methods

III.3.1 *In vitro* Cleavage Assay

For the *in vitro* cleavage assay it was necessary to first isolate DNA from non-transformed tissues. In order to do so, a portion of leaf material that was removed from the seedlings for protoplast isolation was taken, cut into <1mm pieces and DNA isolated using the Qiagen Plant Mini DNA Extraction Kit using the enclosed protocol (Qiagen, Hilden, Germany). This DNA was stored at -20C until usage. In order for digestion

products to be large enough for visual detection, for each cut site a new pair of primers were designed that would give products roughly double (increasing from 180-200 bp to 450-550 bp) in size from the originals (Appendix A). These PCR products were produced using the NEBTaq Polymerase (NEB, Ipswich, MA). The products were then run on agarose gels to verify length and analyzed using a Nanodrop spectrophotometer (ThermoFisher, Waltham, MA). For preparation of RNPs, the designed gRNAs were ordered from Synthego (Synthego, Menlo Park, CA) and Cas9 protein was ordered from IDT (Integrated DNA Technologies, Coralville, IA). The gRNAs were first diluted to 100 µM before being mixed meanwhile the Ca9 was at a concentration of 20 µM and PCR product was at 10 µM. The components were mixed in a ratio of 9:1:1.5 (gRNA:Cas9:PCR product). First 21 µL nuclease free water was mixed with 3 µL NEB 3.1 buffer (NEB, Ipswich, MA), then 1.8 µL of gRNA and 1 µL of Cas9 were added. This was repeated for each individual reaction. After 20 minutes of incubation at 25C, 3 µL of PCR product was added and tubes were relocated to a 37C incubator for 30 minutes for digestion. Afterwards, the whole reaction was added to a 1.6% agarose gel for analysis.

III.3.2 Cas9 Transformation/Expression Detection

To detect the expression of Cas9 *in vivo*, a whole RNA fraction was extracted from transformed protoplasts using the Qiagen Mini Plant RNA Kit (Qiagen, Hilden, Germany) using the enclosed protocol. This RNA fraction, after quantification and purity measurement on NanoDrop spectrophotometer was then stored at -20C until further analysis. The RNA was used in the generation of a cDNA library for Cas9 detection using the Invitrogen SuperScript II cDNA Synthesis Kit (ThermoFisher, Waltham, MA).

Once a cDNA library was available, they were used to generate a PCR product using the NEBTaq Polymerase (NEB, Ipswich, MA). Once completed the product was run on a 1.6% agarose gel for visual confirmation. When a product of desired length was detected, it was extracted using the Qiagen Gel Extraction Kit (Qiagen, Hilden, Germany) for Sanger sequencing.

III.3.3 *In vivo* Nuclease Activity Validation

In order to detect the individual deletion events for *in vivo* Cas9/gRNA activity, PCR products were generated for each of the four cut sites. PCR products were generated from DNA extracted from transformed protoplasts using the NEBTaq Polymerase (NEB, Ipswich, MA). These PCR products were then TA cloned for high resolution sequencing. Using the Invitrogen Original TA Cloning Kit (ThermoFisher, Waltham, MA), the individual PCR products were ligated into the pCR2.1 linearized plasmid. Plasmids were then transformed into heat shock competent DH5α *E. coli*. After clonal selection on ampicillin LB plates, they were mass grown (16 hrs at 37C) in LB liquid broth. The cells were then lysed and plasmids extracted using the Qiagen MiniPrep Kit (Qiagen, Hilden, Germany). Extracted plasmids were then Sanger sequenced using the built in M13 forward primers.

III.4 Results and Discussion

The first validation step in this series of experiments was the *in vitro* cleavage assay. Due to the fact that Cas9 is in a sense a customizable restriction enzyme, it makes sense that one of the quickest ways to validate its sequence specificity is with a traditional

restriction enzyme assay. To begin with this portion, it was vital that uncleaved/untransformed DNA was used. The reason for this being that Cas9 is significantly more sequence specific than per se, a PCR primer [Wu et. al., 2014]. In addition to this site specificity is the consistency in how the nuclease itself digests DNA. In the system used here, the *Streptococcus pyogenes* Cas9, the site of the cut is consistently located 3 base pairs upstream (to the 5') of the PAM sequence. As a result of this, if the DNA template used was one from a previously transformed protoplasts pool then when the RNP attempts to attach and cleave the DNA *in vitro*, there will likely be a mismatch at the location of most importance, namely the cleavage site.

There are a variety of different protocols which list different molar ratios of the components in both mixing the RNPs as well as adding template DNA. The protocol used in this experiment was a fusion of two protocols. The first is the NEB protocol (NEB, Ipswich, MA) which calls for molar ratios of 10:10:1 of gRNA:Cas9:DNA. The other protocol in consideration was provided by Synthego when the gRNA's themselves were delivered which called for molar ratios of 10:1:1 (Synthego, Menlo Park, CA). Ratios for this protocol were then developed by synthesizing the two with consideration given for the relative cost and abundance in our lab of the constituent parts, namely the exorbitant cost of purified Cas9 protein.

Theoretically, the best practice is to have a high proportion of gRNA compared to Cas9 for the simple fact that often for a given experiment, it is best to saturate the enzyme with gRNA to ensure that the highest percentage of nuclease possible is in an RNP configuration. Following this line of logic, it would seem pertinent to saturate the system

with template DNA as well to ensure all RNPs have a target site to act upon. But, if gel electrophoresis is the method of analysis, having too much superfluous DNA can be a detriment in its own rite by creating visual noise like smear and an outshining of cut bands by uncut DNA. Overall the cleavage assay worked well considering it was a fusion of two separate protocols (Figure 3). In order to get the brightest bands (if low effectiveness occurs then bands would be weaker) the whole 30 µL of reaction was added per lane. It is suggested that for future *in vitro* cleavage assays some form of RNAse and proteinase be added post-digestion in order to reduce the amount of noise and give a clearer picture of the products.

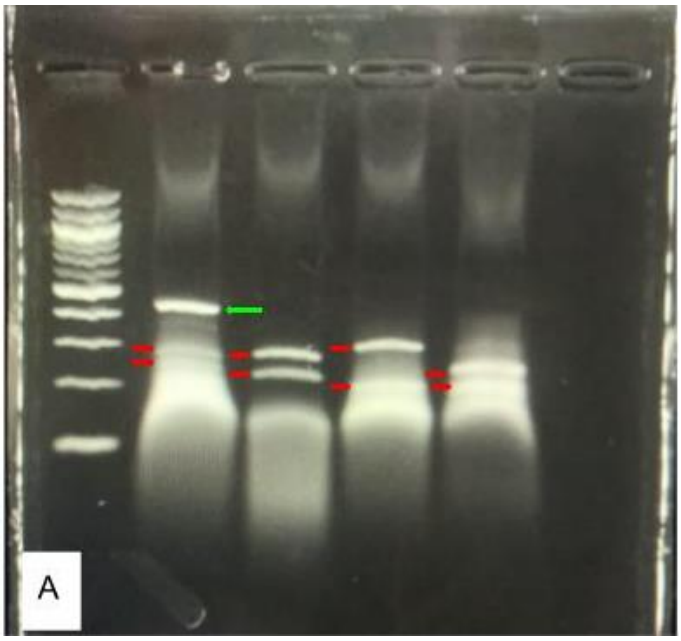


Figure 3: *In vitro* Cas9 RNP digestion gel depicting cut products of four gRNA target sites. Red arrows indicate *in vitro* cleavage products and green arrow shows residual un-cleaved DNA in lane 1. Lanes from left to right are as follows: 1-Region 1 (green arrow 450bp uncut, red arrows 240/235bp products) 2-Region 2 (515bp uncut, red arrows 285/230bp products) 3-Region 3 (522bp uncut, red arrows 313/209bp products) 4-Region 4 (410 bp uncut, red arrows 210/200bp products). The large amount of "smear" is due to the large volume of gRNA's and proteins present.

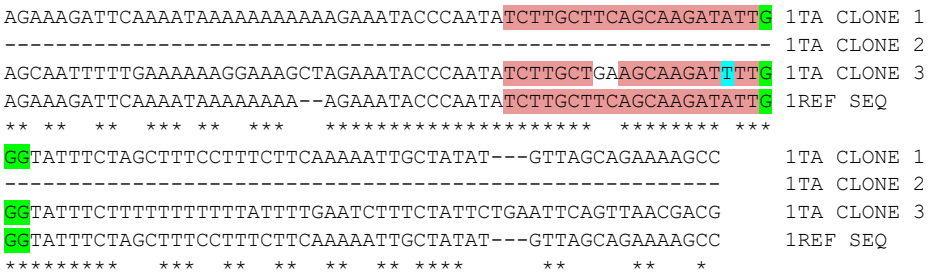
In order to move forward with this information, it was then necessary to see if the plasmid was making its way into the protoplasts, and once their it was necessary to see if the plasmid was in fact being expressed. The simplest way to accomplish both with one step was to look for the expression of the Cas9 protein. Theoretically, by looking simply for the presence of plasmid DNA (considering the variety of wash steps and incubation time would have removed/degraded free plasmids) it could be seen if the plasmid was transfected. But, seeing as expression was as important as transfection, looking for evidence of mRNAs would be more informative. By extracting a whole RNA fraction from the protoplast it's possible to capture a "snapshot" of the metabolic life of the cell in the moment of lysis. As such, if the Cas9 protein was being expressed as expected, the cDNA generated from the Cas9 mRNA would be a perfect candidate for the task. As can be seen in Figure 4, the Cas9 cDNA was first identified via PCR and gel electrophoresis. But in order to bolster the reliability of this, a gel purified product from this band was sanger sequenced. The results proved that the Cas9 protein was in fact being expressed.

As a final validation of the effectiveness of the protocol to this point, it was imperative to look for evidence of Cas9s' nuclease activity *in vivo*. The PCR product that is generated in looking for the deletion was one which could contain both transformed and untransformed sequences, so TA cloning was used for reasons discussed in Chapter IV. Upon closer inspection of the results of the TA cloning/sequencing, several things need to be addressed. A total of 16 TA cloned colonies were selected. The entire first set of four clones sent back results that indicate something went awry (most likely the DNA was sheared by vortexing) in the process of extracting the plasmid and results were discarded. Another set when compared on the MSA aligned, but so poorly and out of the

Figure 4: Multiple Sequence Alignment showing expected and Sanger sequencing output confirming the *in vivo* expression of Cas9 in protoplasts The top row in each segment is the reference sequence for Cas9 of *S. pyogenes*, bottom row is sanger sequencing result. Red indicates a mismatch between sequences and blue indicates homology.

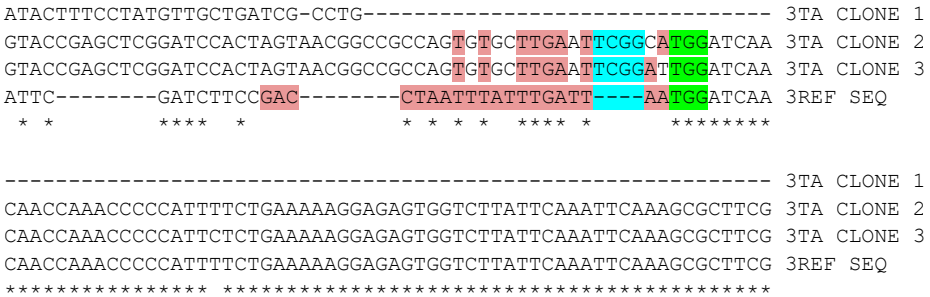
needed frame of reference that their data was nearly useless as well (Figure 5). But within the two data sets for each site that were usable, several anomalies were present. For Region 1, TA clone 3 provided the “cleanest” evidence of a cut with a simple mismatch occurring at the location where it was expected (3 bps 5’ of the PAM sequence). The case for this mismatch being an actual cut is strengthened by the homology present both upstream and downstream of itself. While the possibility of this being a sequencing artifact is present, it seems logical considering the phenomena in question that it is in fact evidence of a cut. Region two is more disappointing. Even though the peaks were clear and homology present throughout the

REGION 1:



REGION 2:

REGION 3:



REGION 4:

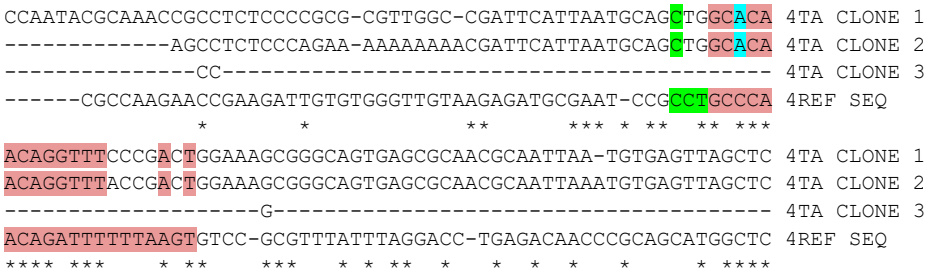


Figure 5: Multiple Sequence Alignment (MSA) analysis of four individual cut sites showing evidence for *in vivo* nuclease activity of Cas9 and designed gRNAs. The regions displayed are of areas proximal to the gRNA location. The red highlight indicates the location of the designed gRNA for each cut, green highlight indicates the PAM sequence to the 3' of each gRNA (region 4 PAM is to 5' due to gRNA being on minus strand). The blue highlights indicate a mismatch or missing base pair where a cut is to be expected. Hashes indicate a region of sequencing data that does not correspond to the rest of the MSA.

gRNA region, no mismatch or gaps were present in any sequencing data provided. Regions 3 and 4 are where the anomalies begin to appear. Region 4 is unique amongst the gRNAs designed in that it is actually designed to operate upon the minus strand, as opposed to the plus strand upon which all other gRNAs were homologous. This meant that the Cas9 in the process of cutting would cut both strands but simply coming from the opposite direction. In this TA clone, the fragment was inserted backwards (due to the directional un-specificity of TA cloning). Upon performing the MSA with the reverse complement of the sequencing FASTA, homology is detected, but only to the 3' direction of the cut. While the homology is not as strong as would be desired, it was definitely stronger in the downstream direction as opposed to the upstream. Again, in the same pattern as observed in Region 1, a mismatch was detected on the third base pair 3' of the PAM (third base pair 5' on the minus strand). While this does on the face indicate a cutting event, the lack of *strong* sequence homology means that it cannot be relied upon as certain evidence for a cutting event. Region 3 is where all the rules seem to go out the window. Downstream (to the 3') of the cut site, *very* strong homology is present for quite some distance (extending another 100bp beyond the sequence displayed). The issues, or rather anomaly, here is what lies to the 5' of the cut site. Here the homology more or less ends, but only between the sequencing data and the reference sequence. In fact, there is still very strong homology in this region between the two sequencing files indicating that *something* was being sequenced and it is not just the result of noise. When BLASTed (blast.ncbi.nlm.nih.gov/Blast.cgi) the homologous sequence to the 5' of the cut site is a nearly perfect match with another region of the rice chromosome. This region on chromosome 11 (12123278-12123353bp) will occur again as an anomaly in Chapter IV.

While this is not evidence of a typical Cas9 deletion *per se*, it does indicate that something is occurring. The same primers used to generate the PCR products for the Region 3 TA cloning were used to generate the original diagnostic sequence that was matched to the reference sequence. Alas, the nature of this is beyond the scope of this chapter and will be further discussed in Chapter IV.

III.5 Conclusion

The aim of this chapter was to discuss the several diagnostic tests used to validate the individual steps in the experimental protocol. In this chapter the evidence is presented that the gRNAs as designed were aptly capable of cutting the regions of DNA they were designed to cut. In addition, it was demonstrated that the transfection/expression stage of the protocol was effective based on the mRNA evidence of Cas9 expression in living protoplasts. Lastly, the evidence for the *in vivo* nuclease activity of the CRISPR construct was presented. While many false starts and anomalies occurred, sequencing data from Region 1 indicated the clearest that *in vivo* nuclease activity was occurring. And with respect to the anomaly present in Region 3, while not hard evidence that the nuclease activity was occurring *as expected*, it indicated that some form of nuclease activity was present by comparing the results to what was experienced with the same PCR/sequencing setup elsewhere in the thesis (Chapter II/IV). And possible causes of this anomaly will be discussed in later chapters.

CHAPTER IV

NUMTs/NUPTs AND LARGE SEGMENT DELETION

IV.1 Synopsis

In this chapter the nature of the NUMTs/NUPTs will be discussed in addition to the bioinformatics work that was used to identify and design targets for large segment deletions. NUclear MiTochondrial (NUMT) insertions and NUclear PlasTid (NUPT) insertions are pieces of the small circular genomes of plastids and mitochondria that have made their way into the nuclear genome through a variety of ways. In this chapter the bioinformatics tools used to identify the insertions as well as the strategies used to develop gRNAs for the removal of a large segment. Lastly, the evidence that could not confirm the large-scale deletions will be discussed.

IV.2 Introduction

Everywhere in the eons of eukaryote evolution certain cellular subunits have played a primary part, particularly in plant proliferation. While their contributions to cellular metabolism as whole entities is more than appreciated today, another key facet of their existence is their tendency to merge with nuclear genomes. This process has been postulated as a way in which new genetic material is integrated for possible evolutionary benefit in certain circumstances [Pantzartzi 2018]. This being proposed because some systems seem to encourage and enable this inter-organellar nucleotide exchange [Munoz-Lopez et al., 2010]. In fact, a fair share of genes in modern higher plants have likely origins in organelles, and in the model plant *Arabidopsis*, up to 4500 genes are of plastid descent

[Noutsos et al., 2005]. But, in recent evolutionary history it is more likely the simple facts of time and the resiliency of plants to damage is the culprit. This is bolstered by the fact that these insertions are nearly ubiquitous in plants [Leister 2005]. The mechanism for the release of this DNA is very simple, sometimes plastids and organelles burst. The mechanism for integration is slightly more complex. As was discussed in the literature review, integration is most likely a byproduct of the natural nuclear processes of DNA repair and transposon migration. As evidence in some species of plants there is a strong correlation between the localization of insertions with transposable elements [Michalovova et al., 2013]. Likewise, the process of genomic repair is active in certain regions like the centromere so it is also no surprise that these integrated elements tend to also accumulate in the pericentric region of chromosomes [Theuri et al., 2005].

While the role of NUPTs and NUMTs as engines of evolution may have fallen to the wayside in recent eons, their usefulness to biology has not [Leister 2005] [Michalovova et. al., 2013]. Most of these regions are of little selective value to cells so they undergo a level of mutation that is relatively consistent with "background mutation" [Sheppard et. al., 2009]. As a result of this these regions are wonderfully useful for phylogenetic studies. By analyzing insertions in relation to plastid genomes and comparing insertions between species they offer key evidence in phylogenetic studies [Michalovova et al., 2013].

The scholarly consensus that these insertions most likely fulfill neither a selective advantage nor housekeeping role means that the deletion of an entire insertion segment would not prove fatal. While it is certain that other large regions of the genome could have been deleted without fatal consequences, the nature of NUMT/NUPTs made them an obvious target for large segment deletion.

Thanks to the variety of different genome browsers and reference sequences, it was possible to find a large insertion that could provide a unique deletion site. Through this chapter it will be discussed how the ideal regions were identified and targeted. These four regions consisted of two nested pairs, a smaller one completely contained within a larger one. These two segments, the larger segment of 107kb (BS) contained within itself a smaller 12.1kb (SS) segment. By designing gRNAs that corresponded to the ends of these large segments, it is possible to then induce a large segment deletion. Large segment deletions have been demonstrated in the past [Zhou et al., 2014]. These large deletions open the door to remove whole families of genes or QTLs of interest. While these deletions have been demonstrated, it has yet to be shown in protoplasts. By opening the door for large deletions in protoplast, it is possible to use this protoplasts system as a screening tool for validation of gRNAs in future large segment deletion work.

IV.3 Materials and Methods

IV.3.1 Bioinformatics

For the preliminary study into the structure of the insertion in question, the MSU rice genome browser (http://rice.plantbiology.msu.edu/) was used. Within this interface, on the 10th chromosome of rice a large insertion was identified. This large insertion spanned from 10807195bp to 10908477bp on the 10th chromosome. The 101.2kb insertion represented a copy of nearly 75% of the chloroplast genome. For this experiment it was desirable to delete the entire chloroplastic insertion. In order to do so the gRNAs for the BS had to be located outside the insertion. To do so the limits of the insertion were identified, then using the NCBI rice genome browser (www.ncbi.nlm.nih.gov/nuccore) it

was possible to search the flanking regions FASTA files for regions of lower repetitiveness that were necessary for PCR primer and gRNA design. Within the NCBI rice genome browser, the Nipponbare reference sequence NC_029265.1 was used [Kawahara et. al., 2013]. By doing so the new large segment (BS) was defined as 10807120-10914170. Within the BS a smaller segment was chosen that had low levels of base pair repetitiveness. Using the same NCBI software and reference sequence, this smaller segment (SS) was defined as 10836909-10849090. From the FASTA data, primer pairs containing each of the four regions were chosen (Figure 6)(Appendix A). As was discussed in Chapter II, the PCR products from these primers were sanger sequenced and used to validate that the reference sequence was homologous to the actual sequence within the plant. This was then used to design gRNAs that corresponded to the four regions/cut sites. As was discussed in Chapter II, these gRNAs were used to perform the transformation of protoplasts from which DNA and RNA fractions were extracted post-incubation.

At this point the method of validation of deletions was flanking PCR. The primers that were used in the *in vitro* cleavage assay were used in this step. For each deletion (BS and SS), a pair of flanking primers was used. These primers produced a product that was then TA cloned. Using the Invitrogen Original TA Cloning Kit (Thermofisher, Waltham, MA), the PCR products were ligated into the pCR2.1 linearized plasmid. After transfection into heat shock competent DH5α *E. coli*, the cells were plated on ampicillin LB selection media. After 24 hours of incubation at 37C, individual colonies were selected and grown in liquid LB media at an ampicillin concentration of 1mL/L. After 18 hours of mass growth in liquid media, the cells were lysed and plasmids harvested with

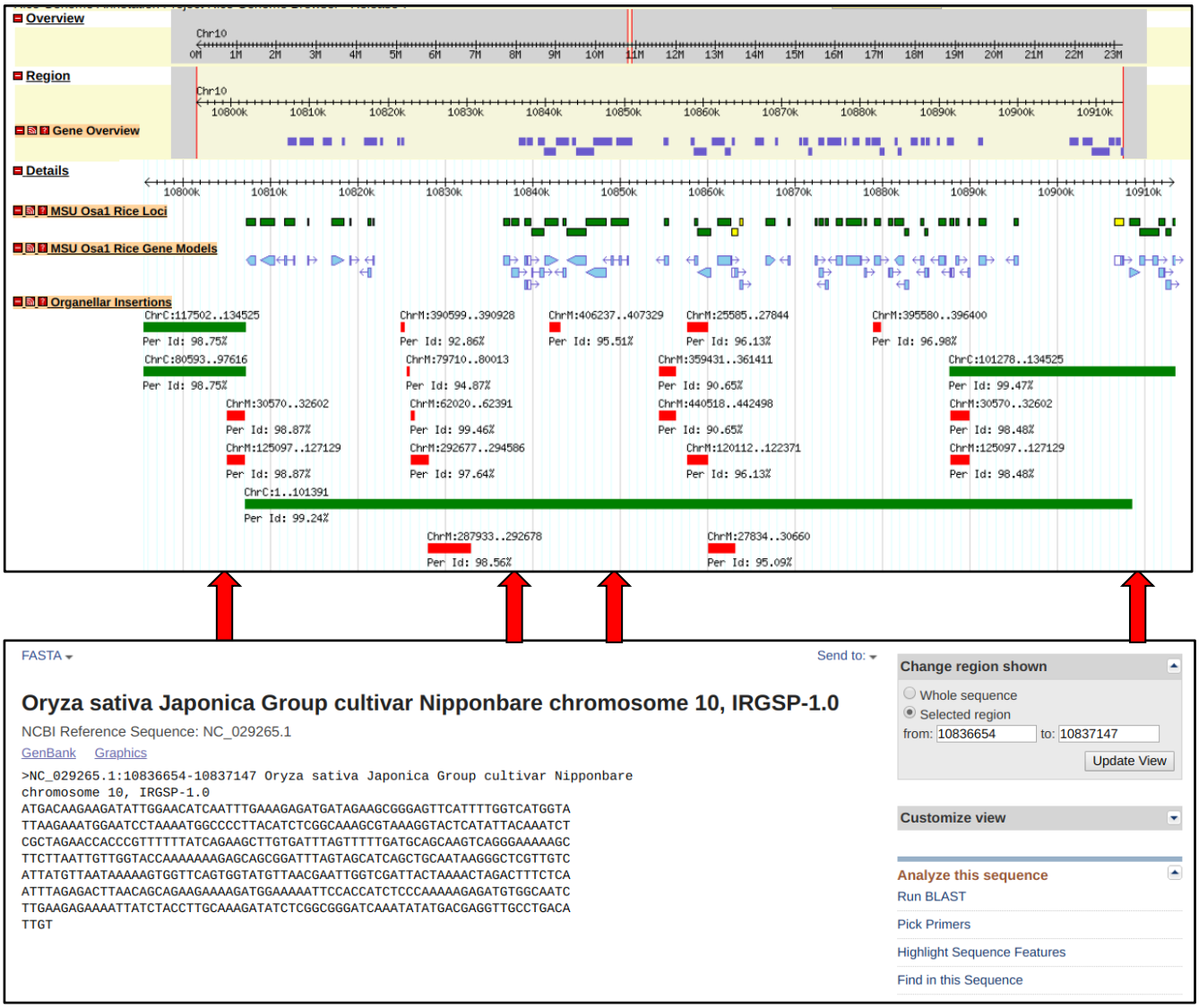


Figure 6: Screenshot of MSU Rice Genome Browser showing diagram of targeted regions. The large green bar across the bottom indicates a chloroplastic insertion ranging from roughly 1080000-10910000bp. The four red arrows indicate the locations of the gRNAs that will be used to induce the large segment (BS) and small segment (SS) deletions. The BS spans arrows 1 and 4, the SS spans arrows 2 and 3. The bottom image is a screenshot of the NCBI rice genome browser that was invaluable in its ability to generate FASTA files of any particular region of interest.

the Qiagen MiniPrep Kit (Qiagen, Hilden, Germany). The plasmids were then quantified on a Nanodrop spectrophotometer (ThermoFisher, Waltham, MA) before sanger sequencing. Once sequencing results were received, the files were analyzed using ClustalOmega MSA (www.ebi.ac.uk/Tools/msa/clustalo).

IV. 4 Results and Discussion

There were a multitude of problems that quickly arose with the deletion of a chloroplast insertion. The chief of which is the fact that the insertions are highly repeated throughout the genome. The BS was unique in that nowhere else in the rice genome did another insertion have the ability to match it. Simply put, the only place in which a deletion event could occur given the two gRNAs for BS is in the intended location. SS was a different story. There were several other locations within the rice genome (chromosomes 4 and 9 had multiple matches) that the two gRNAs could result in a cut. While this is theoretically a non-issue because a deletion anywhere would be a success because it would prove the underlying theory of large deletion events, it does bring into question the accuracy of this method when considering large segment deletions for targeted purposes. Since protoplasts were transformed with a plasmid that contained the gRNAs for *either* BS or SS, not both, and the fact that the other copies of SS throughout the genome have identical flanking regions outside the cut it is difficult to elucidate whether the cuts were happening on the 10th chromosome or somewhere else.

During the process of designing primers to sequence the cut sites another issue was discovered, repetition. Within the insertion there are high levels of repetition, namely repeated genes, that could have confounded results. The remedy to this was simply

moving up or downstream of the intended cut site for unique sequences. The same issue was discovered on a grander scale for the BS flanking sites. Region 4 in particular, was plagued with high sequence repetition. Whether this was true repetition or simply an artifact of its pericentric nature in sequencing is beyond the bounds of this study. But, in order to design a unique sequence for the region 4 primers is was made clear that the site of design would have to be moved several hundred base pairs "downstream" (to the 3') of the insertion increasing the overall size of the intended deletion.

The logic behind the PCR detection mechanism of the deletion is simple. If a deletion had not occurred, the primers would be distal to each other that the kinetics of the PCR reaction would be as if there were only one primer present. In short, nothing would have been produced. Great care was taken to make sure that the primers would not be able to produce an off-target product somewhere else in the genome by screening each pair before delivery. Due to this, the mere fact that a PCR product was present was expected to indicate that a deletion had occurred. But, in accordance with scientific rigor it was necessary to sequence the products.

Here again rises an issue. The nature of the process that is happening means a great deal of variability will occur. The nature of large deletions is that two cuts occur simultaneously enabling the DNA between the cuts to "drift away" before NHEJ occurs [Cai et. al., 2018]. The ensuing repair will most likely result in different products. This is because of two things: degradation of the now loose ends of DNA, and random base pair integration into the annealing site during NHEJ [La Russa et al., 2015]. This variability is exactly what one wants to *avoid* for sanger sequencing. By having a mishmash of products to be sequenced, the peaks of sanger sequencing can become muddled together

degrading the signal and therefore the quality of the read. While gel extraction was initially attempted to rectify this, it was discovered that even this was not enough. Even with size exclusion, different products of the same size can still obscure reads. Through TA cloning it is possible to achieve higher quality reads due to the fact that TA cloning is based on the principal of one PCR product per plasmid, one plasmid per E. coli, one E. coli per colony, and one colony per sequencing input. By utilizing TA cloning it was possible to acquire cleaner reads. Multiple reads were generated from each region (a total of 8 for both the large and small deletions). The expected "anatomy" of these products was to be a whole comprised of two halves; for BS deletion the first half or region 1 and second half of region 4, for SS deletion the first half of region 2 and the second half of region 3 (see Figure 1). When this expected sequence/structure was not discovered it was repeated with each individual sequencing file received. After the same mismatches appeared to be present over and over rather than doubting the sequencing data, the sequencing data was compared to itself. And with some dismay it was realized that all of the sequencing results showed beautiful homology. While this is an accreditation to the sequencing lab in its consistency, it meant that something was wrong on the DNA front. Owing to the fact that all the sequencing files were homologous amongst themselves, it was the same DNA template being provided every time. And this recurring DNA template, was not what was desired.

As can be seen in Figure 7, the results are striking in their apparent randomness as well as their similarities. The matches that correspond to the portion of chromosome 11 are both on the minus strand of the chromosome, whereas the portion of chromosome 1

“Large Deletion” Sequencing Homology

CCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGGCTTGT
AAGCCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTT
GGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACA
CAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACT
CACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCT
GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGC
TTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGG

“Small Deletion” Sequencing Homology

GAATTCGGCTTTCAGTTGCAATAAGGGCTCGACGTGTTGTTTGCTGGGGGATAGATGTCG
CAGGGCAGATCAGATCTCGCATCCACAATCGTTGATCCGTGTGGTACATGGGCCAACAAG
CCACACGAGCTCTCTGTGCGTGTCTGGTACGTACGTCGCGTCCGTGCGATGCAAGGCTAT
AGCGCTTACTCCGGGAAGCCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAG
CTCGGTACCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTA
ATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTAT
TGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCG
AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGC
AGGAAAGAAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCCGTAAAAAAGGCCCG

Figure 7: Sequencing Data depicting the PCR products. Above is a display of a pair of sequences comprised of the homologies between the sequencing files for the “Large Segment Deletion” and “Small Segment Deletion” PCR products. The top sequence, “Large Segment”, contains two overlapping regions. The first (blue/green) is a match to rice chromosome 11 from 12123485-12123278 bp. The second overlap (green/yellow) is a fragment of the Cryl1(A) gene from *Bacillus thuringiensis.* The “Small Deletion” contains the same overlapping regions from the “Large Deletion” on chromosome 11 and Cryl1(A), as well as a portion of chromosome 1 from 23973991-23974151 (red).

contained is on the plus strand. To add to the confusion the gene which is not of rice origin, Cryl1(A), is also the gene which appears to have replaced the front half of Region 3 in the sequencing data from Chapter 3. Both of the chromosomal regions that were represented in the sequence data were of “empty DNA” containing no confirmed or putative genes as well as lacking any evidence of being either NUMTs or NUPTs.

For all of this to be merely a PCR artifact seems highly improbable. The PCR products that were purported to be the evidence of BS deletion and SS deletion were

prepared with DNA from separate transformation events as well as with different pairs of primers. So, the fact that there is so much shared between the two PCR products indicates the same process generating one is generating the other. One of the key assumptions going into this stage of the experiment was that the PCR primer pairs selected were unique to expected deletions. As such, there appears to be little evidence that the PCR primers generated this product. The only place in which there appears to be a match between the PCR primers used and the sequences generated is in one of the sequencing files, but the match is only present in the portion of the sequencing file that was not homologous to the other sequencing files. Aside from that, the PCR primers were not particularly suited to generate the produced products. Out of the 20 base pairs of each primer, a maximum of 10 would match, and not even in sequence at that (meaning a few base pairs would match, followed by a gap, then another match). This level of homology is obviously random. The only way in which it is foreseeable that these poorly matched primers were able to generate products is that even though the matches were poor, they may have been the *best* or rather *least poor* matches possible. But, as with every other assumption made thus far, this one has apparent issues. The primary one being that the different products were produced from different primer pairs. To speculate that each primer pair on its own may have been the *least poor* pair to generate a product is one thing, but to say that somehow the different primer pairs were the *least poor* for both produced products is a stretch. So as far as the bounds of this study are concerned, the reason that these mismatching primer pairs were able to generate *approximately the same product at random* is a mystery. And a puzzling one at that.

The second thing to consider is the apparent chromosomal rearrangement occurring in the “Small Segment” deletion product. In this product, a portion of chromosome 1 on the plus strand is “glued” to the enigmatic chromosome 11/Cryl1(A) fragment which itself is from the minus strand. Now it is known that large scale chromosomal rearrangements can occur with paired cuts, and they can even be targeted/predicted (Shou et. al., 2018). But in order for this to happen then the paired cuts that were predicted for the deletion still occurred, but in an off-target area. In light of this the sequences for the gRNAs were run against the produced products. The pattern here was much the same as with the primers. The gRNAs had some matching base pairs, sometimes in triplets, but again these matching portions were separated by constant and long mismatches. If this was the case and two cuts were occurring simultaneously and a large rearrangement happened, it has to be juxtaposed to the expected outcome. For two cuts to happen simultaneously, it would mean that the gRNAs and Cas9 protein are being expressed at high enough levels that the low probability event that is a dual cut becomes inevitable. If these conditions are met, then it would seem that the obvious result would be the designed cut, not this improbable effect based solely on the criterion of gRNA homology. And even if this improbable rearrangement occurred, it should only be as secondary, less numerous byproducts. Overall the mechanism underlying the apparent chromosomal rearrangement are still a mystery. Not a mystery in mechanism though, because CRISPR off target cuts do occur and NHEJ is rather wily, and for them to happen in a way to produce this product is not *impossible,* merely *very improbable.* No, the mystery here is this: if the conditions for dual cuts and NHEJ are being met, why would an off-target, lower probability event dominate in not one trial, but *all* of them?

So now the obvious question is; is this merely a PCR product that for some reason amplified an existent chromosomal rearrangement? Chromosomal rearrangements come in several flavors, but the flavor that is possible here is a chromosomal translocation. In these events, a piece of a chromosome is transported to another chromosome as a result of a mutation. These occur sporadically in organisms and are quite often fatal. When parsing through the literature there appears to be little homology between the juxtaposed regions on chromosome 11 and 1. In addition, the region of chromosome 11 which was sequenced was unique in its sequence and length to the one site on 11 which has been identified with no other locations matching perfectly throughout the chromosome. Due to the fact that the regions in question do not "overlap" across chromosomes indicates that they are in fact separate chromosomal regions that for whatever reason were included on the same PCR product. The site specificity of the chromosome 11 segment also indicates that the site is not one included within a repetitive element which could have obscured the result even further.

Upon closer examination of the homologies present in the sequencing data it became apparent that a possible cause for all this confusion is the fact that the "Chromosome 11 fragment" is in fact present in the plasmid. In fact, this motif is common amongst many different plasmids to differing degrees. The possible match between this segment and the mysterious *B. thuringiensis* gene is also answered under closer examination. The Cryl1(A) hit is actually the tail end of a vector containing the Cryl1(A) gene which was used for previous study that was logged into the NCBI database. As such, the overlap which appeared to be of Cryl1(A) origin was in fact merely the non-coding portion of a previous experiments vector. All of this is to say that the most likely cause for all of this

confusion is some form of contamination of the plasmid PRGEB32 in the sequencing data. While it could be that the sequence is merely the artifact of the amplification of stray plasmids, it is also possible that the fragment of the plasmid in question was merely acting as its own primer. Since this appears to be the only place within the rice genome that matches a portion of the vector, it could be that the two regions were merely matching up. If the deletion was in fact not occurring and the primers were too far apart to generate a product, the match up between the vector fragment and the chromosome 11 region may have proven to be the best location upon which the polymerase could act.

IV. 5 Conclusion

While the results of this chapter were not the desired ones, strong confirmation of the target deletions, they did provide an interesting insight into the mechanisms of genome editing. The central assumption of the chapter was that the PCR primers designed were unique and would only generate a product if the designed deletion occurred, and it was false. In light of the data presented in previous chapters as well as in this chapter, it appears that the conditions for deletion were met, but the validation process presented was unable to generate a confirmation event of large-scale deletion.

CHAPTER V
SUMMARY AND CONCLUSIONS

The purpose of this thesis was twofold. First, it was intended to design, optimize and then test a protocol for future genome editing work in protoplasts. In this respect it was a success. Healthy protoplast was produced from an often-recalcitrant species, *Oryza sativa,* notably a commercially valuable variety in Texas, Presidio. In addition to this, the process of transforming DNA into the protoplasts was optimized to consistently yield cells that not only contained the targeted DNA but expressed it at noticeable levels. The process of validating gRNA designs in the *in vitro* cleavage assay was proven and optimized for plant DNA extracted via the CTAB method. Lastly the *in vivo* activity of the CRISPR system was validated via an expected result in region 1 in addition to an unexpected result in region 3. As a whole these steps indicate that the process works, and as such it offers an invaluable tool for the future of the CGEL lab in rapidly designing and testing the constructs for future genome editing projects.

Secondly, this thesis aimed at executing and validating a large genome deletion in the protoplast of *Oryza sativa*. In this it failed. While the designed method for validation, the distal flanking PCR assay, was unable to produce a positive result it did bring forth some questions. As was discussed in chapter IV, the mechanisms underlying these strange effects, while having possible *plausible* explanations, remains a mystery.

One thing is certain though, despite the mystery underlying why these products were made, they were made, and made *repeatedly*. The particular motif of the chromosome 11/Cryl1(A) overlap occurred in every “deletion product” sequenced as well as the region

3 cut sequence. Every time a isolation/transformation was performed, the DNA was extracted and these DNA samples (six in total) were used to make the cut site products as well as the "deletion" products.

The weakly held theory from Chapter IV to explain this is as follows. There was a portion of the PREGB32 vector which matched the rice chromosome 11. For whatever reason (most likely a lack of deletion) the primers were unable to produce the desired product. As a result, when the PCR reactions were mixed and exposed to the thermocycler the only place in which the polymerase had a "pseudo-primer" upon which to act was on the location in chromosome 11 paired with the plasmid fragment. This could explain why similar results were produced repeatedly, and in both trials (because the plasmid backbone was the same between large and small segment deletions).

The data presented in Chapters I-III are enticing because they show the different steps of the process (protoplast isolation, transfection, expression, gRNA effectiveness and *in vivo* Cas9 expression/activity) were cooperating. Was there truly no deletion? Or was the PCR based detection process simply too non-specific or error prone to detect it reliably? The suggested future directions are as follows: select a region to delete that shows less repetition both within itself and throughout the chromosome, and attempt other methods for the detection of deletions in an attempt to possibly circumvent the issues that were faced.

As was suggested by a committee member, a strong candidate for detection of deletion events in the future is Loop Sequencing (Loop Genomics, San Jose, CA). In this process the entire genomic DNA sample is partitioned into pieces and each piece is "barcoded" with a unique sequence identifier. These barcoded sequences are then PCR'd

en masse generating a pool of products with each strand of DNA containing an identifier so that the products can then be Illumina sequenced and all reads descending from the "parent strand" can be corroborated. Another approach would simply be to sanger sequence the whole genomic DNA sample. Provided with the forward primer, the sequencer would sequence all of the target strands, both deleted and undeleted. But, evidence for a deletion would be the presence of sequencing data "noise" at and beyond the cut site. If deleted DNA is present in the mix, once that base pair is reached, the uniform signal will begin to be mixed between cut and uncut strands leading to a "static effect" which could potentially identify deletions.

Lastly, a new region of the chromosome, or genome, would be desirable. While chloroplastic insertions are interesting in their various roles, they are simply too numerous. This repetitiveness means that in order to know that your chloroplastic insertion is *your* chloroplastic insertion, and not one of the other 200, you need very wide spacing. In fact, this insertion on chromosome 10 was one of the few viable targets because due to its size, there were no other parts of the chromosome that matched it perfectly. The suggested region for deletion is Chromosome 8, 23326000-23360000. This region contains two separate copies of the amylase gene which is vital for starch degradation within chloroplasts. In addition, this is a relatively unique regions with no major repetitive elements or homologies throughout the genome. In addition, by harvesting seedlings in full daylight and isolating/transforming protoplasts, one could potentially use iodine vapor staining as a physiological marker for successful transformants.

As for protocol itself, namely the protoplasts, it is suggested further optimization occur. While the protoplasts produced in this thesis research were healthy enough to express active Cas9, due to large levels of cellular debris and an obvious lack of high level GFP expression, there is room for improvement. By tweaking the suggested sucrose concentrations for sucrose flotation purification, it may be possible to have a successful sucrose float step in which most of the cellular debris would be removed. In addition, by modifying the concentrations of solutes in the different wash solutions and including an "osmotic balancing period" it may be possible to reduce the levels of stress present on the free-living cells.

In order to ameliorate the issues presented with plasmid/genome interactions it may be a simple fix by introducing a DNase into the protoplasts after the incubation period to digests residual plasmids. If the contamination is occurring from plasmids still within the protoplasts after a DNase treatment it may be pertinent to pursue an RNP based approach. The RNPs proved their utility in the *in vitro* cleavage assay portion of the experiment. By simply carrying these molecules through to the transformation step in lieu of plasmids it may be possible to completely avoid the issues of plasmid/chromosome interactions.

REFERENCES

Bae S, Park J, Kim J-S. 2014. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30: 1473-1475.

Baumgartner BJ, Rapp JC, Mullet JE. 1988. Plastid Transcription Activity and DNA Copy Number Increase Early in Barley Chloroplast Development. Plant Physiol 89: 1011-1018.

Black EM, Giunta S. 2018. Repetitive Fragile Sites: Centromere Satellite DNA as a Source of Genome Instability in Human Diseases. Genes 12: 615-618.

Bottino PJ, Gamborg OL. 1981. Protoplasts in Genetic Modification of Plants. Advances in Biochemical Engineering 19: 239-263**.**

Cai Y, Chen L, Sun S, Wu C, Yao W, Jiang B, Han T, Huo W. 2018. CRISPR/Cas9-Mediated Deletion of Large Genomic Fragments in Soybean. International Journal of Molecular Sciences 19: 3835.

Carroll D. 2017. Genome Editing: Past, Present, and Future. The Yale journal of biology and medicine 90: 653–659.

Cocking E. 1972. Plant Cell Protoplast-Isolation and Development. Annual Reviews in Plant Physiology 23:29-50.

Denbow C, Ehivet SC, Okumoto S. 2018. High Resolution Melting Temperature Analysis to Identify CRISPR/Cas9 Mutants from *Arabidopsis.* BioProtocol 8(14): e2944.

Foster A, Martin-Urdiroz M, Yan X, Wright H, Soanes D, Talbot N. 2018. CRISPR-Cas9 ribonucleoprotein-mediated co-editing and counterselection in the rice blast fungus. Scientific Reports 8: 14355.

Guo X, Ruan S, Hu W, Cai D, Fan L. 2008. Chloroplast DNA Insertions into the Nuclear Genome of Rice: the Genes, Sites and Ages of Insertion Involved. Functional Integrated Genomics 8: 101-108.

Healey A, Furtado A, Cooper T, Henry R. 2014. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods 10: 21-21.

Hofmann R. 2016. A Breakthrough in Monocot Transformation Methods. The Plant cell, *28*(9), 1989. doi:10.1105/tpc.16.00696

Jiang F, Zhu J, Liu HL. 2013. Protoplasts: A Useful Research Tool for Plant Cell Biology, especially dedifferentiation. Protoplasma 250: 1231-1238.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA. 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337: 816-821.

Kawahara Y,Bastide M,Hamilton J.P.,Kanamori H.,et. al. 2013. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data.Rice 6:1-4

Kleinstiver BP, Pattanayak V, Prew MS. 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature:529(7587):490–495.

La Russa MF, Qi LS. 2015. The New State of the Art: Cas9 for Gene Activation and Repression. Molecular and Cellular Biology 35: 3800-3809.

Leister D. 2005. Origin, Evolution and Genetic Effects of Nuclear Insertion of Organelle DNA. Trends in Genetics 21: 655-663.

Liu Y, Vidali L. 2011. Efficient Polyethylene Glycol (PEG) Mediated Transformation of the Moss *Physcomitrilla patens*. Journal of Visualized Experiments 50: 2560.

Matsuo M, Ito Y, Yamauchi R, Obokata J. 2005. The Rice Nuclear Genome Continuously Integrates, Shuffles and Eliminated the Chloroplast Genome to Cause Chloroplast-Nuclear DNA Flux. The Plant Cell 17: 665-675

McClung AM. 2005. Presidio Rice, A New Long Grained Rice with Improved Ratooning Potential and Milling Yield. Texas Rice, Highlighting Research in 2005. p. XI.

Michalovova M, Vyskot B, Kejnovsky E. 2013. Analysis of Plastid and Mitochondrial DNA Insertions in the Nucleus (NUPTs and NUMTs) of Six Plant Species: Size, Relative Age and Chromosomal Localization. Heredity 111: 314-320.

Mojica FJM, Diez-Villasenor A, Garcia-Martines J, Soria E. 2005. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. Journal of Molecular Evolution 60: 174-182.

Munoz-Lopez, Garcia-Perez J. 2010. DNA Transposons: Nature and Applications in Genomics. Current Genomics 11: 115-128.

Naito Y, Hino K, Bono H, Ui-Tei K. 2015. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics 31: 1120-1123.

Noutsos C, Richly E, Leister D. 2005. Generation and Evolutionary Fate of Insertions of Organelle DNA in the Nuclear Genomes of Flowering Plants. Genome Research 15: 616-628.

Pantzartzi C, Pergner J, Kozmik Z. 2018. The role of transposable elements in functional evolution of amphioxus genome: the case of opsin gene family. Scientific Reports 8: 2506.

Portugez S, Martin WF, Hazkani-Covo E. 2018. Mosaic Mitochondrial-Plastid Insertions into the Nuclear Genome Show Evidence of Both Non-Homologous End Joining and Homologous Recombination. BMC Evolutionary Biology 18: 162

Priyadarshani SVGN, Hu B. 2018. Simple Protoplast Isolation System for Gene Expression and Protein Interaction Studies in Pineapple (*Ananas comosus L.*)Plant Methods 14: 95.

Qaim M, Kouser S. 2013. Genetically Modified Crops and Food Security. PLOS One 8: e64879.

Sheppard A, Timmis JN. 2009. Instability of Plastid DNA in the Nuclear Genome. PLOS Genetics 5: e1000323

Shou J, Li J, Liu Y, Wu Q. 2018. Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. Molecular Cell Biology 71: 498-508.

Sun Y, Zhang X, Nie Y, Guo X, Jin S, Liang S. 2004. Production and Characterization of Somatic Hybrids Between Upland Cotton (*Gossypium hirsutum*) and Wild Cotton (*Gossypium klotzshianum Anderss*) via Electrofusion. Theoretical Applied Genetics 109: 472-479.

Theuri J, Phelps-Durr T, Mathews S, Birchlera J. 2005. A Comparative Study of Retrotransposons in the Centromeric Regions of A and B Chromosomes of Maize. Cytogenetic Genome Research 110: 203-208.

Wang D, Qu Z, Adelson D, Zhu JK, Timmis J. 2014. Transcription of Nuclear Organellar DNA in a Model Plant System. Genome Biol Evol 6: 1327-1334

Wu X, Kriz AJ, Sharp PA. 2014. Target specificity of the CRISPR-Cas9 system. Quantitative Biology:2(2):59–70.

Yoshida T, Furihata H, To TK, Kakutani T, Kawabe A. 2019. Genome Defense Against Integrated Organellar DNA Fragments from Plastids into Plant Nuclear Genomes through DNA Methylation. Scientific Reports 9: 2060.

Yoshida T, Hazuka Y, Kawabe A. 2014. Patterns of Genomic Integration of Nuclear Chloroplast DNA Fragments in Plant Species. DNA Research 21: 127-140.

Zallen D. 1993. The “Light” Organism for the Job, Green Algae and Photosynthesis Research. Journal of the History of Biology 26: 269-279.

Zhang K, Deng R, Li Y, Zhang L, Li J. 2016. Cas9 Cleavage Assay for Pre-Screening of sgRNAs Using Nicking Triggered Isothermal Amplification. Chemical Sciences 7: 4951-4957.

Zhang Y, Su J. 2011. A Highly Efficient Rice Green Tissue Protoplast System for the Transient Gene Expression and Studying Light/Chloroplast Related Processes. Plant Methods 7: 30

Zhou H, Liu B. 2014. Large Chromosomal Deletions and Heritable Small Genetic Changes Induced by CRISPR/Cas9 in Rice. Nucleic Acids Research 42: 10903-10914

APPENDIX A

Seedling Growth Media

- .7% agar, 4.4g/L Musharie and Skoog Medium, 8 mM Proline, 1mL/L Plant Preservative Mixture (Plant Cell Technologies)

Protoplast Medias

Wash Solution 1 (WS1)- .5 M Mannitol, 20 mM KCl, 4 mM MES, Stabilize at pH 5.8 with KOH then autoclave

Wash Solution 2 (WS2)- 154 mM NaCl, 125 mM CaCl2, 5 mM KCl, 7 mM MES, Stabilize at pH 5.8 KOH then autoclave

Enzyme Solution (ES)- .4 M Mannitol, 20 mM KCl, 20 mM MES, Stabilize at pH 5.8 then autoclave, for use warm to 55C then add CaCl2 to 10 mM, add BSA to .1%, 15 g/L Cellulase, 4 g/L Macroenzyme. Shake till dissolved.

Mannitol Magnesium Solution (MMG)- .4 M Mannitol, 15 mM MgCl2, 4 mM MES, Stabilize at pH 5.8.

Polyethylene Glycol Solution (PEG)- .2 M Mannitol, .1 M CaCl2, 40% PEG 4000, heat at 55C till dissolved.

Regions and Primers/gRNAs:

Region 1: 10806920-10807500

AGCATTCTACCCGCAATGGTTGGCCATACAATCGCGATTCATAATGGAAAGGAACATATACCTATTTACA
TAACAAATCCTATGGTAGGTCGCAAATTGGGGGAATTCGTACCAACTCGGCATTTCACGAGTTATGAAAG
TGCAAGAAAGGATACTAAATCTCGTCGTTAACTGAATTCAGAATAGAAAGATTCAAAATAAAAAAAAAG
AAATACCCAATATCTTGCTTCAGCAAGATATTGGGTATTTCTAGCTTTCCTTTCTTCAAAAATTGCTAT
ATGTTAGCAGAAAAGCCTTATCCATTAAGAGATGGAACTTCAAGAGCAGCTAGGTCTAGAGGGAAGTTG
TGAGCATTACGTTCGTGCATTACTTCCATACCAAGATTAGCACGGTTGATGATATCAGCCCAAGTATTA
ATAACGCGACCTTGGCT

Forward Primer: GTGCAAGAAAGGATACTAAATCTCG

Reverse Primer: GTTCCATCTCTTAATGGATAAGGCT

Forward Primer Extended: AGCATTCTACCCGCAATGGT

Reverse Primer Extended: GTTGATAGCCAAGGTCGCGT

gRNA: TCTTGCTTCAGCAAGATATT

Region 2: 10836729-10837327

AAATGGCCCCTTACATCTCGGCAAAGCGTAAAGGTACTCATATTACAAATCTCGCTAGAACCACCCGTTTT
TTATCAGAAGCTTGTGATTTAGTTTTTGATGCAGCAAGTCAGGGAAAAAGCTTCTTAATTGTTGGTACCAA
AAAAAGAGCAGCGGATTTAGTAGCATCAGCTGCAATAAGGGCTCGTTGTCATTATGTTAATAAAAAGTGGT
TCAGTGGTATGTTAACGAATTGGTCGATTACTAAAACTAGACTTTCTCAATTTAGAGACTTAACAGCAGAA
GAAAAGATGGAAAAATTCCACCATCTCCCAAAAAGAGATGTGGCAATCTTGAAGAGAAAATTATCTACCTT
GCAAAGATATCTCGGCGGGATCAAATATATGACGAGGTTGCCTGACATTGTGATCGTCCTCGATCAGCAAA
AAGAGTATATAGCTCTTCGGGAATGTGCCATTTTGGGGATTCCTACTATTT
CTTTAGCCGATACAAATTGTGACCCAGATCTCGCGAAT

Forward Primer: TCAGCTGCAATAAGGGCTCG

Reverse Primer: ACAATGTCAGGCAACCTCGT

Forward Primer Extended: AAATGGCCCCTTACATCTCGG

Reverse Primer Extended: ATTCGCGAGATCTGGGTCAC

gRNA: TCAGTGGTATGTTAACGAATT

Region 3: 10848710-10849290

CGATTCCAAATTCCAAGATAACTCATTAGAATTATTAATAAGATGGTCCTGATATATTAGCAATATTTAT
ATTGCCCCTTTTTATTCGCTTTATTACTTCTATTCTAGACCCTATCGTTTATCCTTATGAAATATAATAT
AAATAGAAGGCAGAGGAAAGAGATATAATGAAATTCTTGATTCGATCTTCCGACCTAATTTATTTGATTA
ATGGATCAACAACCAAACCCCCATTTTCTGAAAAAGGAGAGTGGTCTTATTCAAATTCAAAGCGCTTCGT
AATCTTCAACCAGTTCTGTGCTTCAATATAATTTCCCGGAGTAAGCGCTATAGCTTGTTTCCAATACTCA
GCAGCTTGATCAAACCAAGCTTCTGCAATTTCTGAATCACCCTGTAGAATGGCCTGTTCTCCCCGGTCGG
AATAGGTAGTTCCTTCCCCTAGAACCGTACTTGAGAGTTTCCTACCTCATACGGCTCAGAAATTGCTATC
TTAATTTCCCTTGTCTTAACTGAATTCGATT

Forward Primer: TATTCTAGACCCTATCGTTT
Reverse Primer: GCTATAGCGCTTACTCCGGG
Forward Primer Extended: CGATTCCAAATTCCAAGATAACTCA
Reverse Primer Extended: AATCGAATTCAGTTAAGACAAGGGA
gRNA: GACCTAATTTATTTGATTAA

Region 4: 10913870-10914450

CGCCAAGAACCGAAGATTGTGTGGGTTGTAAGAGATGCGAATCCGCCTGCCCAACAGATTTTTTAAGTGTC
CGCGTTTATTTAGGACCTGAGACAACCCGCAGCATGGCTCTATCTTATTGATACGTTACAAAAAATTCCAC
TTGAATCGTCTGATTCCTCTTTATCGAAGAAGCCTGTGCTCAAAATAATCGAGCACGGGCTTTTCTGGTCA
AAACGTATCTTGTCTTTATCACTTTATCATGAGTTCTTTTACTTGGTTAACAATACTTGTTGTTTTGCCGA
TATTTGCGGGTTCA

Forward Primer: AGATTGTGTGGGTTGTAAG
Reverse Primer: TATCGAAGAAGCCTGTGCT
Forward Primer Extended: CGCCAAGAACCGAAGATTGT
Reverse Primer Extended: TGAACCCGCAAATATCGGCA
gRNA: GCCCAACAGATTTTTTAAGT